

УДК 519.2

ПРИМЕНЕНИЕ ЯЗЫКА СТАТИСТИЧЕСКОГО АНАЛИЗА R ДЛЯ АНАЛИЗА ДЕБИТОРСКОЙ ЗАДОЛЖЕННОСТИ ГРАЖДАН ЗА УСЛУГИ ЖКХ (НА ОСНОВЕ ДАННЫХ УПРАВЛЯЮЩЕЙ КОМПАНИИ)

Шубина Д.Е., студент гр. ММ-178, I курс
Научный руководитель: Мешечкин В.В., к.ф.-м.н., доцент
Кемеровский государственный университет
г. Кемерово

Цель данной работы заключается в анализе динамики дебиторской задолженности в сфере жилищно-коммунального хозяйства г. Кемерово при помощи языка программирования R и среды разработки Rstudio [1].

R - это язык программирования, в котором можно писать собственные программы (*скрипты*), а также использовать и создавать специализированные расширения (*пакеты*). Пакет - это набор функций, файлов со справочной информацией и примерами, собранных вместе в одном архиве. Все они проверены на предмет ошибок и находятся в открытом доступе.

Rstudio - интегрированная среда разработки (IDE). По мнению некоторых пользователей и по моему личному опыту, Rstudio имеет более удобный интерфейс, упрощающий работу с R, относительно любой другой среды разработки R.

Прежде чем проводить анализ данных, их необходимо подготовить для работы в R, так как этот язык очень требователен к ним.

На моменте подготовки данных, необходимо собрать как можно больше показателей, характеризующих временной ряд. Однако, в нашей ситуации, имеется в наличии только сам ряд временной задолженности.

Ввиду отсутствия дополнительной информации о показателе задолженности, было принято решение строить прогностическую модель на временном ряду показателя задолженности, но при этом, удалось получить данные до конца 2017 года. Это позволит использовать такие параметры, как частота исследований (периодичность получения данных) в виде 12 месяцев. С 2010 года по 2017 прошло 7 таких циклов.

Начнем построение прогностической модели с анализа временного ряда. Для этого проведем его исследование на наличие сезонности, а так же попробуем сгладить ряд при помощи метода скользящего среднего [2].

Для примера проведем сглаживание временного ряда при помощи скользящего среднего и сделаем это двумя вариантами, с 3 степенью сглаживания [3], при помощи следующих команд:
`>formaSMA3<-SMA(forma, n=3)`

```
>plot.ts(formaSMA3)
```

Результат представлен на рисунке 1.

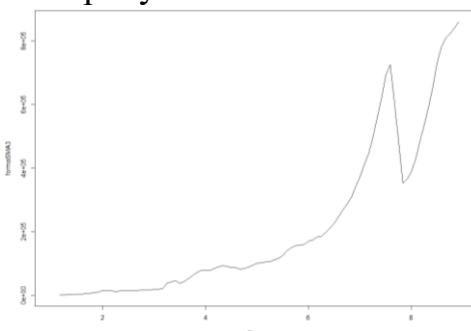


Рисунок 1. Сглаживание 3 порядка

Проведем анализ на периодичность, так как периодические данные состоят из составляющей тренда, периодической составляющей и нерегулярной составляющей. Разложение временного ряда – разделение ряда на эти 3 компоненты, то есть оценка этих составляющих.

Чтобы оценить составляющую тренда и периодическую составляющую, периодического временного ряда, который может быть описан аддитивной моделью, можно использовать функцию `decompose()` [3]. Эта функция оценивает тренд, периодическую и нерегулярную составляющую временного ряда, который может быть описан аддитивной моделью.

Функция `decompose()`:

```
>formaDEC <-decompose(forma)
>plot(formaDEC)
```

Результат выполнения данных инструкция представлен на рисунке 2.

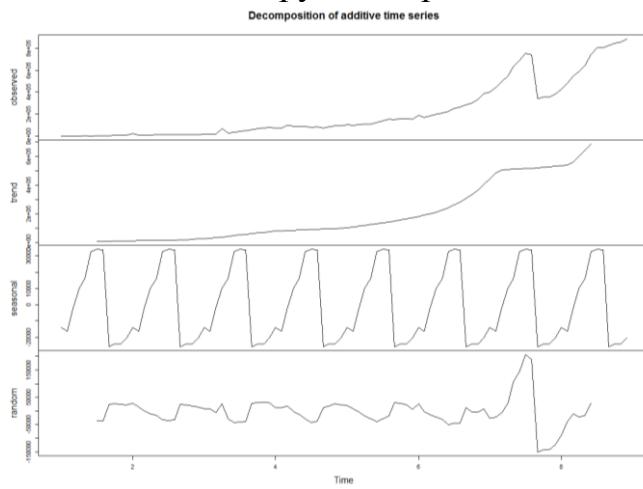


Рисунок 2. Вывод результатов функции `decompose()`

На графиках выше показан исходный временной ряд (верхний график), оценка составляющей тренда (второй сверху), оценка сезонной компоненты (третий сверху) и оценка случайной компоненты (последний). На рисунке 2 видно, что оценка тренда неуклонно растет вверх.

Проведем исключение сезонной компоненты, вычтя оценку сезонной составляющей из исходного временного ряда [3]. Для этого был использован следующий код

```
>formawithoutseasonal<-forma-formaDEC$seasonal  
>plot(formawithoutseasonal)
```

Результат представлен на рисунке 3.

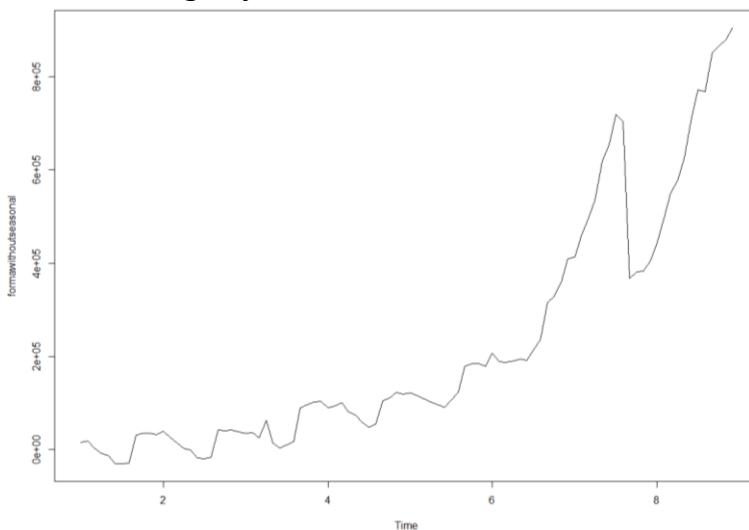


Рисунок 3. Результаты исключение сезонной составляющей

На рисунке 3 видно, что получившийся ряд, не имеет отличий от исходного ряда, следовательно, это подтверждает отсутствие сезонности во временном ряду.

Список литературы:

1. Шубина, Д. Е. Применение статистических методов для анализа дебмторской задолженности граждан за услуги ЖКХ (на основе данных управляющей компании) // Сборник материалов IX всероссийской научно-практической конференции молодых ученых с международным участием «Россия молодая». – Кемерово.: КузГТУ, 2017. – С.774
2. Мешечкин, В. В. Теория прогнозирования / В. В. Мешечкин. – Кемерово, 2016. – 88 с.
3. Роберт, И. R в действии. Анализ и визуализация данных на языке R / И. Роберт – Москва, 2014.